



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

# PRIM1 Deficiency Causes a Distinctive Primordial Dwarfism Syndrome

### Citation for published version:

Parry, DA, Tamayo Orrego, L, Carroll, P, Marsh, JA, Greene, P, Murina, O, Uggenti, C, Leitch, A, The Scottish Genomes Partnership, Káposzta, R, Mer, G, Nagy, A, Orlik, B, Kovács-Pászthy, B, Quigley, AJ, Riszter, M, Rankin, J, Reijns, MAM, Szakszon, K & Jackson, AP 2020, 'PRIM1 Deficiency Causes a Distinctive Primordial Dwarfism Syndrome', *Genes & Development*, vol. 34, pp. 1520-1533.  
<https://doi.org/10.1101/gad.340190.120>

### Digital Object Identifier (DOI):

[10.1101/gad.340190.120](https://doi.org/10.1101/gad.340190.120)

### Link:

[Link to publication record in Edinburgh Research Explorer](#)

### Document Version:

Peer reviewed version

### Published In:

Genes & Development

### General rights

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

### Take down policy

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



# PRIM1 Deficiency Causes a Distinctive Primordial Dwarfism Syndrome

David A. Parry<sup>1+</sup>, Lukas Tamayo-Orrego<sup>1+</sup>, Paula Carroll<sup>1</sup>, Joseph A. Marsh<sup>1</sup>, Philip Greene<sup>1</sup>, Olga Murina<sup>1</sup>, Carolina Uggenti<sup>2</sup>, Andrea Leitch<sup>1</sup>, The Scottish Genomes Partnership, Rita Káposzta<sup>3</sup>, Gabriella Merő<sup>3</sup>, Andrea Nagy<sup>3</sup>, Brigitta Orlik<sup>4</sup>, Balázs Kovács-Pászthy<sup>3</sup>, Alan J. Quigley<sup>5</sup>, Magdolna Riszter<sup>3</sup>, Julia Rankin<sup>6</sup>, Martin A. M. Reijns<sup>1</sup>, Katalin Szakszon<sup>3</sup>, Andrew P. Jackson<sup>1</sup>

1. MRC Human Genetics Unit, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
2. Centre for Genomic & Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK
3. Institute of Pediatrics, Faculty of Medicine, University of Debrecen, Debrecen H-4032, Hungary
4. Institute of Pathology, Faculty of Medicine, University of Debrecen, Debrecen H-4032, Hungary
5. Dept of Radiology, Royal Hospital for Sick Children, Edinburgh EH9 1LF, UK
6. Dept Clinical Genetics, Royal Devon and Exeter NHS Foundation Trust, Exeter EX1 2ED, UK

+ equal contribution

correspondence to APJ, [andrew.jackson@igmm.ed.ac.uk](mailto:andrew.jackson@igmm.ed.ac.uk) and MAMR, [martin.reijns@igmm.ed.ac.uk](mailto:martin.reijns@igmm.ed.ac.uk)

## Abstract

DNA replication is fundamental for cell proliferation in all organisms. Nonetheless, components of the replisome have been implicated in human disease and here we report *PRIM1*, encoding the catalytic subunit of DNA primase, as a novel disease gene. Using a variant-classification agnostic approach, biallelic mutations in *PRIM1* were identified in five individuals. *PRIM1* protein levels were markedly reduced in patient cells, accompanied by replication fork asymmetry, increased inter-origin distances, replication stress and prolonged S-phase duration. Consequently, cell proliferation was markedly impaired, explaining the patients' extreme growth failure. Notably, phenotypic features distinct from those previously reported with DNA polymerase genes were evident, highlighting differing developmental requirements for this core replisome component that warrant future investigation.

## Introduction

Microcephalic primordial dwarfism (MPD) disorders constitute several Mendelian syndromes characterized by intrauterine growth retardation, short stature, and microcephaly (Klingseisen and Jackson 2011). Previously, partial loss of function mutations have been identified in genes encoding fundamental components of the cell-cycle machinery, suggesting a model in which disrupted cell proliferation during development results in “hypocellular” forms of dwarfism (Klingseisen and Jackson 2011).

Perturbation of components of the DNA replication machinery has emerged as a common cause of MPD, involving several stages of the replication process. In the G1-phase of the cell cycle the pre-replication complex acts to license replication origins, and subsequently the replication pre-initiation complex is formed, with the full replisome assembled at S-phase entry, to initiate DNA replication effected by three DNA polymerases, alpha, delta and epsilon (Pol  $\alpha$ ,  $\delta$  and  $\epsilon$ ). Biallelic mutations in components of the pre-replication complex were first identified, with *ORC1*, *ORC4*, *ORC6*, *CDT1*, and *CDC6*, causing Meier-Gorlin Syndrome (MGS) (Bicknell et al. 2011b, 2011a; Guernsey et al. 2011), a disorder defined by a

triad of microtia, patella hypoplasia and growth restriction. Heterozygous stabilizing mutations in *GMNN* and recessive variants in *CDC45*, *MCM5* and *DONSON* have subsequently been associated with MGS (Burrage et al. 2015; Vetro et al. 2017; Fenwick et al. 2016; Reynolds et al. 2017). Mutations in pre-initiation complex components *MCM4*, and *GIN51* cause distinct microcephalic dwarfism syndromes with immunodeficiency and/or adrenal failure (Gineau et al. 2012; Cottineau et al. 2017). Finally, mutations in genes encoding subunits of the replicative DNA polymerases *POLA1*, *POLD1*, *POLD2*, *POLE* and *POLE2* have implicated components of the active replisome in primordial dwarfism, often with immune deficiency, and in the case of *POLE*, adrenal failure (Pachlopnik Schmid et al. 2012; Thiffault et al. 2015; Frugoni et al. 2016; Logan et al. 2018; Van Esch et al. 2019; Conde et al. 2019).

Each of these genes has been shown to be essential for cellular survival in large-scale functional screens (Table S1) (Chen et al. 2017) while constraint metrics from the gnomAD consortium suggest that loss of a single copy is unlikely to result in developmental disease (Table S2) (Karczewski et al. 2019). Recessive pathogenic variants are therefore most likely to lead to perturbation of function more severe than haploinsufficiency but falling short of biallelic loss of function. Consistent with this, previously published variants associated with recessive MPD have included hypomorphic missense or splice altering variants that result in at least one allele with residual function. The recent identification of an intronic splice altering mutation in *POLE* in 12 families exemplifies this, where each affected individual inherited different loss of function alleles in *trans* with the same intronic variant (c.1686+32C>G) causing a hypomorphic splicing defect. Hypomorphic alleles such as this may not be annotated as deleterious by variant classifiers, so we considered whether additional genes for MPD could be detected by using a variant-classification agnostic approach to identify as yet undiscovered cryptic mutations in essential genes. Here, we report the identification of *PRIM1* as an MPD gene, and demonstrate that mutations reduce cellular *PRIM1* protein levels, impairing DNA replication.

## Results

### Identification of Biallelic Variants in *PRIM1*

We analyzed whole genome (WGS) and whole exome (WES) sequence data from 220 families with microcephalic dwarfism spectrum disorders (occipital-frontal circumference [OFC]  $\leq -4$  SD; height  $\leq -2$  SD), to find variants enriched in this cohort relative to the general population and inherited in a pattern consistent with recessive inheritance (Materials & Methods, Supplemental Table S3). This identified three families sharing a homozygous intronic variant (c.638+36C>G, Fig. 1A,B) in *PRIM1*, which encodes the catalytic subunit of DNA primase.

Supporting its pathogenicity, the intronic *PRIM1* variant was significantly enriched in our microcephalic dwarfism cohort in comparison to the Genome Aggregation Database where just 2 heterozygotes out of 141,456 individuals were observed (gnomAD v2.1.1,  $p=3.61 \times 10^{-15}$ , all populations, combined WES and WGS data, Supplemental Table S4).

Inbreeding coefficients in all three families were consistent with parental relatedness (Supplemental Table S5). As well, P1-P4 had large overlapping regions of homozygosity between 10.4 Mbp and 32.7 Mbp in length on chromosome 12 containing the *PRIM1* gene (Supplemental Fig. S1). These findings provided additional genetic support for *PRIM1* being the causative gene as they were consistent with the expectation for consanguineous families with a rare recessive disease that a disease-causing variant is expected to lie within a region of homozygosity in each family (Lander and Botstein 1987). Within the shared region of homozygosity there was also a common 2 Mb haplotype surrounding the *PRIM1* variant in all 4 affected individuals (Supplemental Fig. S1, Supplemental Table S7). This indicates that all 3 families share a distant common ancestor (Supplemental Fig S1, Supplemental Table S5), establishing the causal mutation to be within the haplotype and excluding the only other enriched variant in families 2 and 3, a missense change in *OR6C4*.

Analysis of the c.638+36C>G variant with SpliceAI (Jaganathan et al. 2019), NNSplice (Reese 1997) and MaxEntScan (Yeo and Burge 2004) predicted activation of a cryptic donor site, resulting in the inclusion of 31 nucleotides of intronic sequence following exon 6, expected to disrupt the open reading frame. This, alongside PRIM1 function in DNA replication, provided further support for disruption of PRIM1 causing MPD.

To identify additional disease-causing *PRIM1* alleles, we searched exome data from a subset of 149 trios and singletons from the Deciphering Developmental Disorders (DDD) project selected on the basis of morphometric criteria (OFC  $\leq$  -3 SD and height  $\leq$  -3 SD) and absence of previous diagnostic variants identified by DDD. This identified a single individual (P5) compound heterozygous for an essential splice donor variant (c.103+1G>T) and a missense variant (c.901T>C, p.Cys301Arg, hereafter referred to as C301R). These variants were respectively extremely rare or absent from public databases (Table 1). Additional analysis of exome data from P5 did not reveal any other likely diagnostic variants as an alternative explanation. The c.103+1G>T variant results in loss of the donor site of intron 1 while the c.901T>C variant alters a cysteine residue, conserved in vertebrates, to a physiochemically dissimilar arginine. This substitution was predicted damaging by *in silico* tools including PolyPhen2, SIFT, MutationTaster and CADD (CADDv1.6 Phred score=28.4).

### **PRIM1 subjects share a distinct clinical phenotype**

A similar clinical phenotype was evident in all five individuals. All had extreme growth failure, with *in utero* growth retardation (birth weight  $-3.4 \pm 0.6$  SD) and severe postnatal growth restriction (height  $-8.8 \pm 2.0$  SD) (Fig. 1C, Supplemental Table S8). While relatively macrocephalic, in absolute terms there was extreme microcephaly (OFC  $-6.0 \pm 1.5$  SD). On neuroimaging this was reflected by a ‘microcephaly with simplified gyri’ appearance (Fig. 1D). A common facial appearance was apparent, with prominent forehead and triangular face, with blepharophimosis  $\pm$  microphthalmia, micrognathia and small low-set ears (Fig. 1E). Absence of subcutaneous fat, and distally tapered fingers were noted in all affected individuals on clinical examination. Hypothyroidism was frequent alongside significant

haematological/immune dysfunction. All had hypo/agammaglobulinemia (Table 2). In four of five cases there was persistent lymphopenia accompanied by intermittent anemia/thrombocytopenia. Documented episodes of sepsis occurred in several patients. In addition, episodes of fever without an identified infective source were recorded. All individuals had chronic respiratory symptoms, and four died in early childhood from either respiratory or GI infections. Hepatic dysfunction was noted clinically, and at postmortem, hepatic fibrosis, cirrhosis or macronodular regeneration was evident in three of these individuals, suggesting preceding chronic liver inflammation.

The distinct clinical phenotype in these individuals was consistent with a shared genetic etiology, taken together with our genetic data, led us to conclude that biallelic mutations in the *PRIM1* gene cause MPD. Next, we sought to test the functional impact of the variants identified in *PRIM1*.

### **The c.638+36C>G variant activates a cryptic splice site reducing cellular PRIM1 protein**

To confirm that the c.638+36C>G variant resulted in mis-splicing, RT-PCR was performed on RNA extracted from a lymphoblastoid cell line (LCL) established from individual P2 and from primary fibroblasts cultured from individual P3. This demonstrated the presence of an additional PCR product in P2 and P3 corresponding in size to the predicted mis-spliced transcript (Fig. 2A). Sanger sequencing of these PCR products confirmed *in silico* splice predictions, with the presence of an additional 31 nucleotides of intronic sequence (Fig. 2B,C) leading to substitution of a lysine to serine at codon 214 and a frameshift with premature termination 21 amino-acids later (HGVS nomenclature: p.Lys214Serfs\*21, referred to hereafter as K214fs). This would be anticipated to lead to nonsense-mediated decay (NMD) of this transcript, with any remaining translated truncated protein not containing key catalytic residues and therefore enzymatically inactive (Fig. 1B). Reduced cellular levels of full-length protein were therefore expected, and consistent with this, marked reduction in full-length PRIM1 protein levels was evident on immunoblotting of cell

extracts from P2 and P3 in comparison to control cell lines (Fig. 2D; 14% of wild-type levels in patient LCL and 9% in fibroblasts). (A band at ~25 kDa, that might correspond to a non-functional truncated PRIM1 protein (27 kDa), was seen inconsistently in blots from P2 LCL extracts but not in P3 fibroblasts, Supplemental Fig. S2).

### **C301R and c.103+1G>T variants also reduce PRIM1 protein levels**

P5 was compound heterozygous for c.103+1G>T and C301R variants. However, a patient cell line was not available, therefore the consequences of these variants were assessed by different methodologies. Our initial expectation was that essential splice site mutation would be a null allele and the missense variant, partial loss of function.

C301 is a residue in a buried hydrophobic region of the protein (Fig. 3A), and its substitution with a bulky, positively charged arginine residue (C301R) was predicted to be highly destabilizing by FoldX (Schymkowitz et al. 2005) (Supplemental Fig. S3). We therefore assessed the effect of this substitution on PRIM1 protein levels, developing a FACS-based assay to measure the level of PRIM1-GFP containing this insertion relative to a co-translated RFP control (Fig 3B). This approach corrects for cell to cell variation in transfection efficiency and ensures that assessment of protein expression is not confounded by transcriptional differences or mRNA stability. This assay confirmed substantially reduced PRIM1 protein levels for the C301R substitution ( $25 \pm 0.3\%$ , Fig. 3C,D).

We also modelled this mutation in budding yeast, where substitution of the equivalent residue in *S. cerevisiae* Pri1p to arginine (L309R) also led to a reduction in protein level (Supplemental Fig. S4), whereas the wild-type human cysteine residue (L309C) had no effect on either the level of Pri1 protein (Supplemental Fig. S4) or growth (Supplemental Fig. S5). Similar degradation dynamics of wild-type and L309R Pri1p in cycloheximide chase experiments (Supplemental Fig. S4) suggest that reduced protein levels are most likely the consequence of incorrect folding of the C301R/L309R protein rather than an increase in turnover. Notably C301 is in close proximity to the key catalytic residue R304 (Kirk and



Kuchta 1999), suggesting that it might not solely act to reduce protein levels. Consistent with this, the L309R mutation was lethal in yeast (Supplemental Fig. S5), raising the possibility that C301R in human PRIM1 is also a null allele, as disturbance of this catalytic site could potentially lead to loss of function in any mutant protein that persists in the cell.

The C301R substitution appeared to be of greater functional severity than expected, so we examined the c.103+1G>T variant in more detail, considering whether it might instead be hypomorphic in effect. Consistent with this possibility, a minigene assay established that the c.103+1G>T variant could lead to the use of a nearby cryptic splice site (Fig. 3E, Supplemental Fig. S6) resulting in a 3 amino acid in-frame insertion in the PRIM1 transcript ('V35insDGV', Fig. 3F,G). The FACS-based stability assay established that the V35insDGV insertion also lowers PRIM1 protein levels ( $52 \pm 0.5\%$  of wild-type levels, Fig. 3C,D), but less severely than C301R, consistent with this being a partial loss of function mutation.

In conclusion, these experiments confirmed that both variants have functional impact on the PRIM1 protein, that in combination would still leave residual PRIM1 activity.

### **Fork stability and origin firing is impaired in PRIM1 primary fibroblasts**

Human *PRIM1* encodes the 49 kDa catalytic subunit of the DNA primase heterodimer, which forms a heterotetrameric complex with the two DNA polymerase  $\alpha$  subunits (Czechowska and Blasiak 2005). PRIM1 is responsible for the synthesis of short RNA primers required for the initiation of DNA replication and Okazaki fragment synthesis, before handover of these primers to Pol  $\alpha$  for DNA synthesis (Rowen and Kornberg 1978; Frick and Richardson 2001). As such, depletion of PRIM1 would be expected to impact the efficiency of DNA replication.

Doubling times of P3 primary fibroblasts were significantly increased compared to controls (Fig. 4A, P3 =  $49 \pm 3.6$  hours, C1 =  $23 \pm 2.9$  hours, C2 =  $27 \pm 2.3$  hours,  $P < 0.001$ ) (Fig. 4A), and S-phase length, measured by sequential CldU/IdU pulse labelling (Martynoga et al. 2005), was substantially increased (P3 =  $16.6 \pm 0.12$  hours, C1 =  $6.1 \pm 0.4$  hours, C2 =  $7.2 \pm 1$  hours) (Fig. 4B,C). Complementation by transient transfection of wild-type, but not C301R

mutant PRIM1 confirmed that the lengthened S-phase was specifically due to PRIM1 deficiency in patient cells (Fig. 4D,E). Impaired efficiency of DNA replication was also observed in BrdU-DNA content flow cytometry which demonstrated reduced BrdU incorporation in P3 primary fibroblasts during S-phase compared with control fibroblast lines ( $P < 0.001$ , Fig. 4F,G). In addition,  $\gamma$ H2AX levels, a marker of DNA damage, were significantly increased in patient fibroblasts during S-phase (Fig. 4H), as were levels of chromatin bound RPA (Supplemental Fig. S7), consistent with the presence of endogenous replication stress. However, whether S-phase checkpoint activation occurs remains to be determined.

FACS analysis of the cell cycle also demonstrated an increased proportion of G1-phase cells (Fig. 4F, Supplemental Fig. S8), without significant differences in the fraction of cleaved caspase-3 or p21-positive patient cells (Supplemental Fig. S9). We therefore concluded that the hypoproliferative phenotype was most likely the consequence of delayed cell-cycle progression rather than increased apoptosis or cell-cycle exit.

We used DNA-fiber combing to further characterize replication defects in PRIM1 deficient patient cells. Primary fibroblasts from P3 demonstrated similar replication defects to fibroblasts from an individual with POLE-IMAGe syndrome caused by Pol  $\epsilon$  deficiency (P1 from Logan et al. 2018). Fork speed was not affected in either PRIM1 or POLE cell lines (Fig. 5B) but both cell lines displayed a similar, high degree of perturbed fork-progression compared to controls (Fig. 5C), indicative of elevated fork stalling and restart events, as might be expected for an enzyme continually required for priming of Okazaki fragments during lagging strand synthesis. Increased inter-origin distances were also observed [median IOD 152 Kbp vs. 113 Kbp,  $P < 0.001$ ], (Fig 5D), suggesting reduced DNA replication initiation at origins. These alterations in replication dynamics were similarly observed in P2 LCLs (Supplemental Fig. S10).

Taken together, these findings were consistent with impaired replication fork formation and increased replication fork stalling in PRIM1 deficient cells.

## Discussion

We report here the identification of five individuals with biallelic deleterious variants in *PRIM1*, using a variant-classification independent approach. *PRIM1* is essential to cellular survival (Chen et al. 2017) and, like other replisome components, knockout in mice is lethal during embryogenesis (Blake et al. 2011). Mutations in essential genes which result in developmental disorders are necessarily hypomorphic alleles. Identification of such variants is challenging, particularly as they may often be in non-coding regions (Tarnauskaitė et al. 2019; Cottineau et al. 2017; Logan et al. 2018). Consequently, they may not appear on variant lists filtered by functional consequence. Therefore, the agnostic approach used in this study may have utility for the identification of further genes for other Mendelian diseases.

*PRIM1* deficiency results in similar growth restriction to other replication-associated disorders such as those associated with Pol  $\alpha$  and Pol  $\epsilon$  deficiency (Supplemental Fig. S11, Supplemental Fig. S12). Microtia and lymphopenia are also reminiscent of other replisome disorders (Conde et al. 2019; Logan et al. 2018). Likewise, reduced subcutaneous fat in *PRIM1* deficient individuals parallels the lipodystrophy found in *POLD1* cases (who had normal growth) due to heterozygous single-codon deletion of the catalytic site of Pol  $\delta$  (Weedon et al. 2013). However, there are distinctive clinical features for individuals with *PRIM1* deficiency, particularly early childhood mortality occurring in four out of five cases. While this might be accounted for by the significant B cell lymphopenia and accompanying hypo/agammaglobinemia, other disease processes may have also contributed. Chronic lung disease, recurrent diarrhea and liver dysfunction could also reflect a persisting inflammatory state. Notably a specific intronic variant in *POLA1* causing an X-linked reticulate pigmentary disorder, is also associated with recurrent pneumonias and chronic diarrhea, with underlying immunodeficiency alongside autoinflammation (Starokadomskyy et al. 2016). However, the *POLA1* associated reticulate pigmentary skin phenotype is not reported in any of the individuals with *PRIM1* variants identified herein. Most individuals with MPD caused by *POLA1* variants have growth retardation and microcephaly without such features (Van

Esch et al. 2019), and facially *PRIM1* deficiency appears distinct from these and individuals with *POLE*-associated IMAGE syndrome (Logan et al. 2018). Likewise, blepharophimosis, microphthalmia, hypothyroidism, episodic thrombocytopenia and anemia do not appear to be features reported in other replisome disorders.

In conclusion, we describe a distinct form of microcephalic dwarfism associated with *PRIM1* mutations. Similar to *POLE*/A1 DNA polymerase disorders, replication fork defects are seen in *PRIM1* deficient cells. This likely reflects an inability of these cells to establish sufficient replication origins and sustain lagging strand synthesis at a sufficient rate to meet the demand for rapid replication. The ensuing delay to the cell cycle and reduced cellular proliferation provides a likely mechanism for the resulting “hypocellular” microcephalic dwarfism. Given the necessity of several polymerases to act in concert in every replicating cell, the different phenotypic manifestations of Pol  $\alpha$ , Pol  $\delta$ , Pol  $\epsilon$  and DNA primase related disorders are surprising and require future investigation.

## Materials and Methods

### Research subjects

Genomic DNA was extracted from blood samples by standard methods or from saliva samples using Oragene collection kits according to the manufacturer's instructions. Informed consent was obtained from all participating families. This study was approved by the Scottish Multicentre Research Ethics Committee (05/MRE00/74). Parents provided written consent for the publication of photographs of the affected individuals.

The DDD study has UK Research Ethics Committee approval (10/H0305/83, granted by the Cambridge South REC, and GEN/284/12 granted by the Republic of Ireland REC).

### Whole-Genome Sequencing and Variant Calling

Whole Genome Sequencing in P1, P2 and P4 was performed by Edinburgh Genomics (Clinical Division) as part of the Scottish Genomes Project (SGP) to 30X coverage using TruSeq Nano library preparation kits and a HiSeq X sequencing platform (Illumina). FASTQs generated by Edinburgh Genomics were aligned to the human genome (hg38, including alt, decoy and HLA sequences) using bwa mem(Li 2013) (0.7.13). Post-processing was performed with samblaster (Faust and Hall 2014) (0.1.22) to mark duplicate reads, and the Genome Analysis ToolKit(McKenna et al. 2010) (GATK, v3.4-0-g7e26428) for indel realignment and base recalibration. Genotype likelihoods for each sample were calculated using the GATK HaplotypeCaller and resulting GVCF files were called jointly using GATK's GenotypeGVCFs function. Variant quality score recalibration (VQSR) was performed as per GATK best-practices(Van der Auwera et al. 2002) and a truth sensitivity threshold of 99.9% applied. Functional annotations were added using Ensembl's Variant Effect Predictor (McLaren et al. 2016) (v90).

### Variant Filtering and Identification of a Variant Activating a Cryptic Splice

#### Donor in *PRIM1*

Cohort WGS variant data were filtered to only include variants with an allele frequency  $< 0.5$  % in all gnomAD populations at biallelic sites outside of repeat masked or low-complexity regions using VASE (v0.1, <https://github.com/david-a-parry/vase>). Remaining variants were analyzed against gnomAD v2.1 data converted to hg38 coordinates.

Where variants were present in gnomAD WES and WGS VCFs, allele counts per population were extracted from VCFs. For variants absent from either or both gnomAD WES or WGS VCFs, homozygous reference allele counts were derived for the respective samples using sequencing depth data, where the fraction of individuals with sequencing depth  $\geq 5$  were counted as homozygous reference. Variants were annotated with P-values from one-tailed Fisher's exact tests comparing the alternate/reference allele counts in non-related affected individuals in the MPD cohort versus individual population groupings in gnomAD.

Annotated variants were assessed for segregation consistent with recessive inheritance. At least one allele was required to have P-values below  $5 \times 10^{-5}$  for each of the African/African-American, Latino/Admixed American, East Asian, Finnish, Non-Finnish European and South Asian gnomAD subpopulations. The second allele for a gene was required to either match these same criteria or to be a predicted loss-of-function variant. Analysis was restricted to canonical transcripts and biallelic combinations of alleles observed in non-affected family members were excluded. Where available, phase information from parental genotypes and/or physical phasing was used to exclude biallelic combinations in *cis*.

This enrichment strategy could be confounded by differing variant calling sensitivity in gnomAD compared to the MPD cohort and by population stratification. In order to mitigate these effects we required enriched alleles to be below the given P-value threshold in each of the population groups listed above and then ranked genes based on incidental occurrences of variants within the MPD cohort, gene essentiality and the maximum (least significant) observed P-value. Incidental occurrences were defined as the number of families where an individual was a carrier for a variant yet no qualifying second allele was observed in the affected individual(s). Additionally, more than 2 qualifying alleles in the same family were

also defined as incidental occurrences to address the presence of ‘noisy’ genes. Gene essentiality was defined as the fraction of studies in OGEE (Chen et al. 2017) in which a gene was found to be cell-essential. A ranking score for variants was calculated as follows:

$$s = (1 + oi)(1 + fi)(1.1 - ess)$$

Where “*oi*” indicates the number of incidental occurrences in other families, “*fi*” indicates the number of incidental occurrences within the same family and “*ess*” represents the fraction of studies in which the gene is considered cell essential. A lower score was considered “better” such that variants were penalised for incidental occurrences and lack of evidence for gene essentiality.

These methods are implemented in a python package (ReVERSE), available at <https://git.ecdf.ed.ac.uk/dparry/reverse>.

### **Segregation Filtering of DDD CAPS14 Cohort**

VASE (v0.1, <https://github.com/david-a-parry/vase>) was used to perform variant filtering and segregation analysis. Variants overlapping *PRIM1* were filtered to remove those with a frequency greater than 0.5% in gnomAD or dbSNP151. Additionally, those variants not predicted to either be protein altering variants, within a splice region or with a SpliceAI (Jaganathan et al. 2019) delta score  $\geq 0.5$  were also excluded. Genotype calls were filtered if samples had fewer than 5 reads covering the site or the fraction of variant reads to total reads (VAF) were less than 0.25. Remaining variants were assessed for familial segregation consistent with recessive inheritance.

Following identification of *PRIM1* compound heterozygous variants in F5, variant data for this family was re-analysed to test exome-wide for any other possible causative variants. Data for this family was analysed exome-wide for recessive variants as above and also analysed for de novo variants, requiring a VAF  $\geq 0.25$  in the proband and VAFs  $< 0.05$  in both parental genotypes and total site depths of  $\geq 8$  in parents and proband.

## Plasmid construction

All oligonucleotides and plasmids used in this study are listed in Supplemental Tables S9 and S10 respectively. Plasmids pMAR780 and 781 were generated by cloning annealed oligonucleotides into *BclI*/*SwaI* digested pML104 and pML107 plasmids, respectively (Laughery et al. 2015). To construct pMAR782, the *PRI1* (*YIROO8C*) coding sequence, without stop codon and with a silent PAM site mutation (c.1945G>C), was amplified by PCR using MRY201 genomic DNA as template and cloned into *XbaI*/*BamHI* digested pGFP-C-FUS (Niedenthal et al. 1996). pMAR782 allows expression of C-terminally GFP-tagged Pri1 protein in medium lacking methionine. To construct pMAR790, a 1.35 kb fragment of human genomic DNA (chr12:56750975-56752323, hg38) covering *PRIM1* exon1 and exon 2 was PCR amplified and cloned into *BspEI*/*XbaI* sites of RHCglo (a gift from Thomas Cooper; Addgene plasmid # 80169) (Singh and Cooper 2006), fusing *PRIM1* exon 1 with exon 1 of the minigene construct. The donor (c.103+1G>T) splice site mutation was introduced in intron 1 using site directed mutagenesis (NEB Q5 Site-Directed Mutagenesis Kit), generating pMAR793. Human C-terminally EGFP-tagged PRIM1 was synthesized by IDT flanked by *NheI* and *BspEI* sites and provided in pUCIDT-AMP GoldenGate (pMAR796). The *NheI*/*BspEI* fragment was then transferred into pmGFP-P2A-KO-P2A-RFP (a gift from Ramanujan Hegde; Addgene plasmid # 105686) (Juszkiewicz and Hegde 2017), replacing mGFP and creating pMAR797. To introduce C301R and VDGins mutations, Q5 SDM was performed on pMAR796, creating pMAR798 and 802 respectively, and the sequence-verified PRIM1-EGFP *NheI*/*BspEI* fragments cloned into pmGFP-P2A-KO-P2A-RFP to make pMAR799 and 803. All plasmids were confirmed by restriction digest analysis and Sanger sequencing.

## Cells and Cell culture

Primary dermal fibroblasts were established from skin punch biopsies and maintained in AmnioMAX medium (Thermo Fisher Scientific, Cat. No. 17001074) in 5% CO<sub>2</sub> and 3% O<sub>2</sub>. Patient Lymphoblastoid cell lines (LCLs) were generated in house from peripheral blood



samples by EBV transformation using standard methods. LCLs were maintained in RPMI 1640 supplemented with 15% FBS, L-glutamine and penicillin/streptomycin antibiotics. HeLa (ATCC) cells were maintained in DMEM supplemented with 10% FBS and 5% penicillin/streptomycin in 5% CO<sub>2</sub>, under normoxic conditions. Genotypes of patient cell lines were validated by Sanger sequencing.

### **RT-PCR**

Total RNA was extracted from cell lines using the RNeasy Mini Kit (Qiagen) according to the manufacturer's instructions. Following DNase I (Qiagen) treatment cDNA was generated using SuperScript III Reverse Transcriptase (Thermo Fisher Scientific). RT-PCR was performed on cDNA using primers in exon 4 and 8 (Supplemental Table S9).

### **Minigene splicing assay**

RPE1 cells were seeded (4 x 10<sup>5</sup> cells/well) in a 6-well plate. After 24 h, cells were transfected with 800 ng of pMAR790 or pMAR791 plasmid in Opti-MEM Reduced Serum Medium with Lipofectamine 2000 according to the manufacturer's instructions. Cells were harvested after 24 h, RNA extracted using the RNeasy Mini Kit (Qiagen), and cDNA generated using SuperScript™ III Reverse Transcriptase (Invitrogen). PCR amplification was then performed with RSV5U and RTRHC primers, and wild-type and mutant cDNA amplicons resolved by agarose gel electrophoresis to visualize splicing differences. PCR products were cloned into pGEM-T Easy (Promega), and Sanger sequencing performed on 20 clones.

### **Immunoblotting**

Whole cell extracts were prepared by lysis using 50 mM Tris-HCl (pH 8), 280 mM NaCl, 0.5% NP40, 0.2 mM EGTA, 1 mM DTT and 10% glycerol, supplemented with Roche Complete Protease Inhibitor Cocktail. Protein samples were resolved on 4-12% NuPAGE Bis-Tris gels (Thermo Fisher Scientific) and transferred to nitrocellulose or PVDF membranes. Subsequent immunoblotting was performed using the following antibodies: anti-p48 Primase (8G10; Cell Signalling, 4725, 1:1,000), anti-Vinculin (hVIN-1; Sigma-Aldrich,

V9264, 1:2,000), anti- $\alpha$ -Tubulin (B-5-1-2; Sigma-Aldrich, T6074, 1:10,000), anti-GFP (Living Colors JL-8; Clontech, 1:4,000), anti-FLAG (M2; Sigma-Aldrich, F1804, 1:2,000) or anti-Actin (Sigma-Aldrich A2066, 1:4,000).

Finally, detection was performed using Amersham ECL Prime Western Blotting Detection Reagent on the ImageQuantLAS4000 (GE Healthcare Life Sciences), or the Odyssey CLx Imaging System (LI-COR Biosciences). Quantifications were performed using ImageQuant TL 7.0 and Image Studio Lite 5.2, respectively.

### **FACS-based Dual Reporter Stability Assay**

C-terminally EGFP-tagged PRIM1 (wild-type or mutants) was cloned into a mammalian expression vector as described under Plasmid Construction. Expression from the CMV promoter of the resulting constructs (pMAR797 and derivatives) produces mRNA encoding PRIM1-EGFP-P2A-FLAG-SR-P2A-RFP, which when translated due to intervening P2A “self-cleaving” peptide sequences produces 3 individual polypeptides in equimolar amounts: PRIM1-EGFP, FLAG-SR and RFP. This allows comparison of the relative level of PRIM1-EGFP (wild-type and mutants) and RFP in individual cells.

RPE1 cells were seeded ( $4 \times 10^5$  cells/well) in a 6-well plate. and a day later transfected with 2  $\mu$ g of the reporter plasmid in Opti-MEM Reduced Serum Medium with Lipofectamine 2000 as per manufacturer’s instructions. At 24 h post-transfection live cells were assayed for the presence of GFP and RFP using a BD Biosciences LSR Fortessa flow cytometer and data analyzed using FlowJo software (v7.6.1, Tree Star).

### **BrdU-DNA content Flow cytometry**

Primary Fibroblasts were seeded into AmnioMAX medium (Life Technologies) to achieve ~60% confluency after 24 h. Cells were then incubation with 10  $\mu$ M BrdU for 30 min, washed, harvested and fixed with 70% EtOH at -20°C for 16 h. Fixed cells were then digested with 1 mg/ml pepsin, denatured in 2 M HCl for 15 min, and washed with PBS. After blocking in 0.5% BSA, 0.5% Tween-20, BrdU labelling was detected using anti- BrdU antibody

(Abcam, ab6326, 1:75) and anti-Rat Alexa-Fluor 488 secondary antibody (Thermo Fisher Scientific, A11006). DNA content was determined by co-staining with 50 µg/ml propidium iodide. Cells were assayed on a BD Biosciences LSR Fortessa flow cytometer and data analysed using FlowJo software (v7.6.1, Tree Star).

### **Cell Proliferation Doubling Times**

Primary fibroblasts ( $1.5 \times 10^5$  cells) were seeded on day 0 into a T25 flask. Cell counts were performed every 3 days using a MOXI Z Mini automated cell counter (MXZ001). After counting,  $1.5 \times 10^5$  cells were reseeded into a new T25. Doubling times were calculated during log-phase growth (day 3 to day 12) using the formula:  $t / \log_2(\frac{e}{b})$  where  $t$  = time in hours,  $e$  = final population size and  $b$  = population size at the start of log-phase growth.

### **S-phase time measurements and PRIM1 complementation (CldU/IdU Pulse Labelling)**

S-phase time ( $T_s$ ) measurements were performed using sequential pulses of CldU and IdU to determine the fraction of cells that leave S-phase in a specified time window, according to the formula  $T_s = T_i \cdot S_{\text{cells}} / L_{\text{cells}}$ , where  $T_i$  represents the time (hours) between pulses,  $S_{\text{cells}}$  are the number of cells that are in S-phase at the first pulse (all CldU positive), and  $L_{\text{cells}}$  are the number of cells that have left S-phase at the time of the second pulse (CldU positive and IdU negative cells)..

Primary fibroblasts, cultured on coverslips, were pulsed with 25 µM CldU for 1.5 h followed by a 30 min 125 µM IdU pulse. Cells were then fixed in 4% PFA.

For complementation experiments, P3 fibroblasts were electroporated with dual expression vectors (pMAR797 and derivatives: empty vector or containing GFP-PRIM1 and GFP-PRIM1-C301R) using the Invitrogen Neon Kit (Cat. No. MPK10096) according to manufacturer instructions (2.5 µg plasmid DNA for  $1 \times 10^6$  cells, 20 ms pulse width). 24 h after electroporation, cells were pulsed with CldU and IdU (as above), trypsinised and sorted on a BD FACS Aria to select GFP-positive, successfully transfected cells. Cells were collected

in AmnioMAX C-100, seeded on coverslips in 24-well plates for 1.5 h and subsequently fixed in 4% PFA. Immunofluorescence for CldU and IdU as outlined below, was performed using Rat anti-BrdU (Abcam, Cat. No. ab6326) and mouse anti-BrdU (BD, Cat. No. 347580).

### **Immunofluorescence**

Fibroblasts were grown on coverslips for 24 h, fixed with 4% PFA, washed and blocked for 1 h in 10% goat serum/PBS/0.1% Triton. Independent coverslips were incubated with p21 (1:200, Calbiochem Cat. No. OP64), cleaved caspase 3 (1:1000, Cell Signaling, Cat. No. 9661),  $\gamma$ H2AX (1:1000, Millipore, Cat. N. 05-636) or RPA2 (1:200, Calbiochem, Cat. No. NA19L antibodies. After overnight incubation in primary antibodies, signals were detected using fluorescently conjugated antibodies and co-stained with DAPI. For all antibodies excepting cleaved caspase 3 and p21, cells were pre-extracted for 5 min on ice with ice-cold buffer (25 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM EDTA, 3 mM MgCl<sub>2</sub>, 300 mM sucrose and 0.5% Triton X-100) before fixation to remove soluble proteins and detection of chromatin-bound fractions. For  $\gamma$ -H2AX staining, cells were pulsed with 20  $\mu$ M EdU (Sigma, Cat No. 900584) for 1 h before fixation. EdU was detected subsequent to immunofluorescence using the click reaction and Azide-Alexa Fluor-488 (ThermoFisher A10266).

### **DNA Combing**

Primary fibroblast cells exponentially growing in AmnioMAX C-100 (Thermo Fisher Scientific, Cat. No. 17001074) medium under hypoxic conditions were pulsed with 25 mM CldU (Sigma, Cat. No. C6891) for 20 min, washed with pre-warmed PBS and then pulsed with 125 mM IdU (Sigma, Cat. No. I7125) for 20 min. After trypsinization,  $5 \times 10^5$  cells were used to cast 3-4 agarose (Biorad, Cat. No. 1613111) plugs per condition and processed for DNA combing according to a previously described protocol (Gallo et al. 2016), omitting the SCE buffer plug digestion steps. IdU and CldU were detected using mouse anti-BrdU (BD, Cat. No. 347580) and rat anti-BrdU (Abcam, Cat. No. ab6326), respectively. DNA was

detected using anti-ssDNA antibody (Millipore, MAB 3034). Images were acquired on a widefield microscope (Zeiss Axiophot) with a 63X or 40X lens. The 2.33 Kbp/ $\mu\text{m}$  elongation rate ( $\mu\text{m}$  to Kbp conversion) was obtained from bacteriophage lambda DNA combing and measurement, as previously described (Gallo et al. 2016). Measurements and analysis were performed using ImageJ. DNA fork speed was obtained dividing the length of the IdU tracks adjacent CldU tracks (ongoing forks) by the IdU incubation time (20 min) and expressed in Kbp/min. Fork asymmetry is presented as left IdU vs. right IdU ratios. Inter-origin distances (IODs) correspond to the space (in Kbp) between the centre points of adjacent bi-directional replication origins.

## **Consortia**

Members of the Scottish Genome Partnership include Timothy J. Aitman, Andrew V. Biankin, Susanna L. Cooke, Wendy Inglis Humphrey, Sancha Martin, Lynne Mennie, Alison Meynert, Zosia Miedzybrodzka, Fiona Murphy, Craig Nourse, Javier Santoyo-Lopez, Colin A. Semple, and Nicola Williams.

## **Acknowledgements**

We thank the families and clinicians for their involvement and participation; John Wyrick, Jean Beggs, Thomas Cooper, Ramanujan Hegde, Aziz El Hage for reagents; the IGMM Advanced Imaging Resource for advice; E. Freyer for assistance with FACS analysis; IGMM core sequencing service; Ailith Ewing for helpful discussion; and Edinburgh Genomics (Clinical Division) for WGS sequencing. This work was supported by the European Union's Horizon 2020 research and innovation program ERC Advanced Grant (grant agreement 788093); by a UK Medical Research Council Human Genetics Unit core grant (MRC, U127580972), and the Scottish Genomes Partnership. LT-O is also supported by an Embo ALTF 739-2019. J.A.M. is supported by an MRC Career Development Award (MR/M02122X/1) and is a Lister Institute Research Prize Fellow. The Scottish Genomes Partnership is funded by the Chief Scientist Office of the Scottish Government Health Directorates (SGP/1) and The Medical Research Council Whole Genome Sequencing for Health and Wealth Initiative. The DDD Study presents independent research commissioned by the Health Innovation Challenge Fund [grant number HICF-1009-003]. This study makes use of DECIPHER (<http://decipher.sanger.ac.uk>), which is funded by Wellcome. See Nature PMID: 25533962 or [www.ddduk.org/access.html](http://www.ddduk.org/access.html) for full acknowledgement.

## **Author Contributions**

APJ and DAP designed the study. DAP analyzed genetic data. LT-O, PC, MAMR, OM, CU and AL performed laboratory experiments. LT-O, PC, MAMR, DAP and APJ analyzed experimental data. JAM and DAP analyzed protein structure data. RK, GM, AN, BO, BK-P, AJQ, MR, JR, KS and APJ recruited and evaluated study participants. DAP, APJ, LT-O and MAMR wrote the manuscript.

## References

- Bicknell LS, Bongers EMHF, Leitch A, Brown S, Schoots J, Harley ME, Aftimos S, Al-Aama JY, Bober M, Brown PAJ, et al. 2011a. Mutations in the pre-replication complex cause Meier-Gorlin syndrome. *Nat Genet* **43**: 356–359.
- Bicknell LS, Walker S, Klingseisen A, Stiff T, Leitch A, Kerzendorfer C, Martin C-A, Yeyati P, Al Sanna N, Bober M, et al. 2011b. Mutations in ORC1, encoding the largest subunit of the origin recognition complex, cause microcephalic primordial dwarfism resembling Meier-Gorlin syndrome. *Nat Genet* **43**: 350–355.
- Blake JA, Bult CJ, Kadin JA, Richardson JE, Eppig JT. 2011. The Mouse Genome Database (MGD): premier model organism resource for mammalian genomics and genetics. *Nucleic Acids Res* **39**: D842–8.
- Burrage LC, Charng WL, Eldomery MK, Willer JR, Davis EE, Lugtenberg D, Zhu W, Leduc MS, Akdemir ZC, Azamian M, et al. 2015. De Novo GMNN Mutations Cause Autosomal-Dominant Primordial Dwarfism Associated with Meier-Gorlin Syndrome. *Am J Hum Genet* **97**: 904–913.
- Chen W-H, Lu G, Chen X, Zhao X-M, Bork P. 2017. OGEE v2: an update of the online gene essentiality database with special focus on differentially essential genes in human cancer cell lines. *Nucleic Acids Res* **45**: D940–D944.
- Conde CD, Petronczki ÖY, Baris S, Willmann KL, Girardi E, Salzer E, Weitzer S, Ardy RC, Krolo A, Ijspeert H, et al. 2019. Polymerase  $\delta$  deficiency causes syndromic immunodeficiency with replicative stress. *J Clin Invest* **129**: 4194–4206.
- Cottineau J, Kottemann MC, Lach FP, Kang YH, Vély F, Deenick EK, Lazarov T, Gineau L, Wang Y, Farina A, et al. 2017. Inherited GINS1 deficiency underlies growth retardation along with neutropenia and NK cell deficiency. *J Clin Invest* **127**: 1991–2006.
- Crooks GE, Hon G, Chandonia JM, Brenner SE. 2004. WebLogo: A sequence logo generator. *Genome Res* **14**: 1188–1190.
- Cunningham F, Amode MR, Barrell D, Beal K, Billis K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fitzgerald S, et al. 2015. Ensembl 2015. *Nucleic Acids Res* **43**: D662–9.
- Czechowska A, Błasiak J. 2005. Eukaryotic DNA polymerases. *Postepy Biochem* **51**: 130–139.
- Faust GG, Hall IM. 2014. SAMBLASTER: fast duplicate marking and structural variant read extraction. *Bioinformatics* **30**: 2503–2505.
- Fenwick AL, Kliszczak M, Cooper F, Murray J, Sanchez-Pulido L, Twigg SRF, Goriely A, McGowan SJ, Miller KA, Taylor IB, et al. 2016. Mutations in CDC45, Encoding an Essential Component of the Pre-initiation Complex, Cause Meier-Gorlin Syndrome and Craniosynostosis. *Am J Hum Genet* **99**: 125–138.

- Frick DN, Richardson CC. 2001. DNA Primases. *Annu Rev Biochem* **70**: 39–80.
- Frugoni F, Dobbs K, Felgentreff K, Aldhekri H, Al Saud BK, Arnaout R, Ali AA, Abhyankar A, Alroqi F, Giliani S, et al. 2016. A novel mutation in the POLE2 gene causing combined immunodeficiency. *J Allergy Clin Immunol* **137**: 635-638.e1.
- Gallo D, Wang G, Yip CM, Brown GW. 2016. Analysis of replicating yeast chromosomes by DNA combing. *Cold Spring Harb Protoc* **2016**: 138–149.
- Gineau L, Cognet C, Kara N, Lach FP, Dunne J, Veturi U, Picard C, Trouillet C, Eidenschenk C, Aoufouchi S, et al. 2012. Partial MCM4 deficiency in patients with growth retardation, adrenal insufficiency, and natural killer cell deficiency. *J Clin Invest* **122**: 821–32.
- Guernsey DL, Matsuoka M, Jiang H, Evans S, MacGillivray C, Nightingale M, Perry S, Ferguson M, Leblanc M, Paquette J, et al. 2011. Mutations in origin recognition complex gene ORC4 cause Meier-Gorlin syndrome. *Nat Genet* **43**: 360–365.
- Jaganathan K, Kyriazopoulou Panagiotopoulou S, McRae JF, Darbandi SF, Knowles D, Li YI, Kosmicki JA, Arbelaez J, Cui W, Schwartz GB, et al. 2019. Predicting Splicing from Primary Sequence with Deep Learning. *Cell* **176**: 535-548.e24.
- Juszkiewicz S, Hegde RS. 2017. Initiation of Quality Control during Poly(A) Translation Requires Site-Specific Ribosome Ubiquitination. *Mol Cell* **65**: 743-750.e4.
- Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alföldi J, Wang Q, Collins RL, Laricchia KM, Ganna A, Birnbaum DP, et al. 2019. Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. *bioRxiv* 531210.
- Kirk BW, Kuchta RD. 1999. Arg304 of human DNA primase is a key contributor to catalysis and NTP binding: Primase and the family X polymerases share significant sequence homology. *Biochemistry* **38**: 7727–7736.
- Klingseisen A, Jackson AP. 2011. Mechanisms and pathways of growth failure in primordial dwarfism. *Genes Dev* **25**: 2011–2024.
- Laughery MF, Hunter T, Brown A, Hoopes J, Ostbye T, Shumaker T, Wyrick JJ. 2015. New vectors for simple and streamlined CRISPR-Cas9 genome editing in *Saccharomyces cerevisiae*. *Yeast* **32**: 711–720.
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.
- Logan C V., Murray JE, Parry DA, Robertson A, Bellelli R, Tarnauskaitė Ž, Challis R, Cleal L, Borel V, Fluteau A, et al. 2018. DNA Polymerase Epsilon Deficiency Causes IMAGE Syndrome with Variable Immunodeficiency. *Am J Hum Genet* **103**: 1038–1044.
- Martynoga B, Morrison H, Price DJ, Mason JO. 2005. Foxg1 is required for specification of ventral telencephalon and region-specific regulation of dorsal telencephalic precursor proliferation and apoptosis. *Dev Biol* **283**: 113–127.



- McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, Garimella K, Altshuler D, Gabriel S, Daly M, et al. 2010. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res* **20**: 1297–1303.
- McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GRS, Thormann A, Flicek P, Cunningham F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol* **17**: 122.
- Niedenthal RK, Riles L, Johnston M, Hegemann JH. 1996. Green fluorescent protein as a marker for gene expression and subcellular localization in budding yeast. *Yeast* **12**: 773–86.
- Pachlopnik Schmid J, Lemoine R, Nehme N, Cormier-Daire V, Revy P, Debeurme F, Debré M, Nitschke P, Bole-Feysot C, Legeai-Mallet L, et al. 2012. Polymerase  $\epsilon$ 1 mutation in a human syndrome with facial dysmorphism, immunodeficiency, livedo, and short stature ("FILS syndrome"). *J Exp Med* **209**: 2323–30.
- Reese MG. 1997. Improved splice site detection in Genie. In *Journal of Computational Biology*, Vol. 4 of, pp. 311–323, Mary Ann Liebert Inc.
- Reynolds JJ, Bicknell LS, Carroll P, Higgs MR, Shaheen R, Murray JE, Papadopoulos DK, Leitch A, Murina O, Tarnauskaitė Ž, et al. 2017. Mutations in DONSON disrupt replication fork stability and cause microcephalic dwarfism. *Nat Genet* **49**: 537–549.
- Rowen L, Kornberg A. 1978. Primase, the dnaG protein of Escherichia coli. An enzyme which starts DNA chains. *J Biol Chem* **253**: 758–764.
- Singh G, Cooper TA. 2006. Minigene reporter for identification and analysis of cis elements and trans factors affecting pre-mRNA splicing. *Biotechniques* **41**: 177–181.
- Starokadomskyy P, Gemelli T, Rios JJ, Xing C, Wang RC, Li H, Pokatayev V, Dozmorov I, Khan S, Miyata N, et al. 2016. DNA polymerase- $\alpha$  regulates the activation of type I interferons through cytosolic RNA:DNA synthesis. *Nat Immunol* **17**: 495–504.
- Tarnauskaitė Ž, Bicknell LS, Marsh JA, Murray JE, Parry DA, Logan C V., Bober MB, Silva DC, Duker AL, Sillence D, et al. 2019. Biallelic variants in *DNA2* cause microcephalic primordial dwarfism. *Hum Mutat* **40**: humu.23776.
- Thiffault I, Saunders C, Jenkins J, Raje N, Canty K, Sharma M, Grote L, Welsh HI, Farrow E, Twist G, et al. 2015. A patient with polymerase E1 deficiency (POLE1): clinical features and overlap with DNA breakage/instability syndromes. *BMC Med Genet* **16**: 31.
- Van der Auwera GA, Carneiro MO, Hartl C, Poplin R, del Angel G, Levy-Moonshine A, Jordan T, Shakir K, Roazen D, Thibault J, et al. 2002. From FastQ Data to High-Confidence Variant Calls: The Genome Analysis Toolkit Best Practices Pipeline. John Wiley & Sons, Inc.
- Van Esch H, Colnaghi R, Freson K, Starokadomskyy P, Zankl A, Backx L, Abramowicz I, Outwin E, Rohena L, Faulkner C, et al. 2019. Defective DNA Polymerase  $\alpha$ -Primase

Leads to X-Linked Intellectual Disability Associated with Severe Growth Retardation, Microcephaly, and Hypogonadism. *Am J Hum Genet* **104**: 957–967.

Vetro A, Savasta S, Russo Raucci A, Cerqua C, Sartori G, Limongelli I, Forlino A, Maruelli S, Perucca P, Vergani D, et al. 2017. MCM5: a new actor in the link between DNA replication and Meier-Gorlin syndrome. *Eur J Hum Genet* **25**: 646–650.

Weedon MN, Ellard S, Prindle MJ, Caswell R, Allen HL, Oram R, Godbole K, Yajnik CS, Sbraccia P, Novelli G, et al. 2013. An in-frame deletion at the polymerase active site of POLD1 causes a multisystem disorder with lipodystrophy. *Nat Genet* **45**: 947–950.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. In *Journal of Computational Biology*, Vol. 11 of, pp. 377–394.

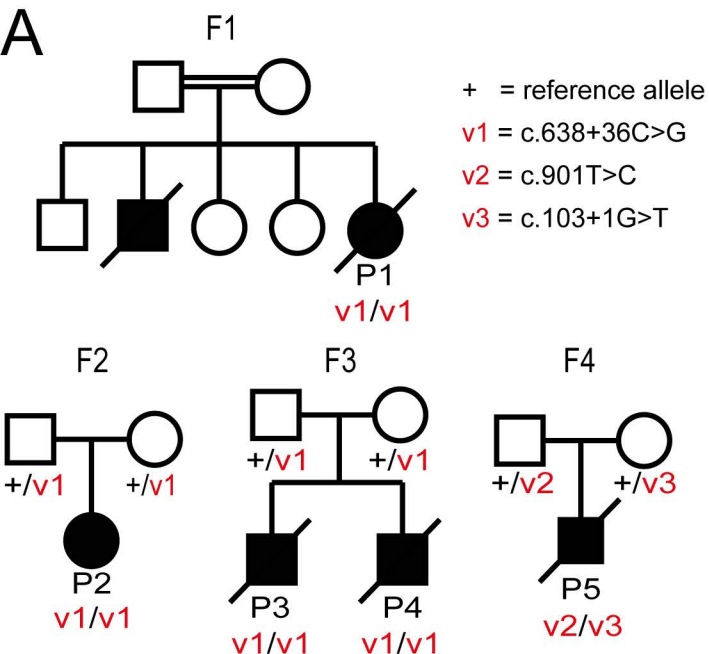
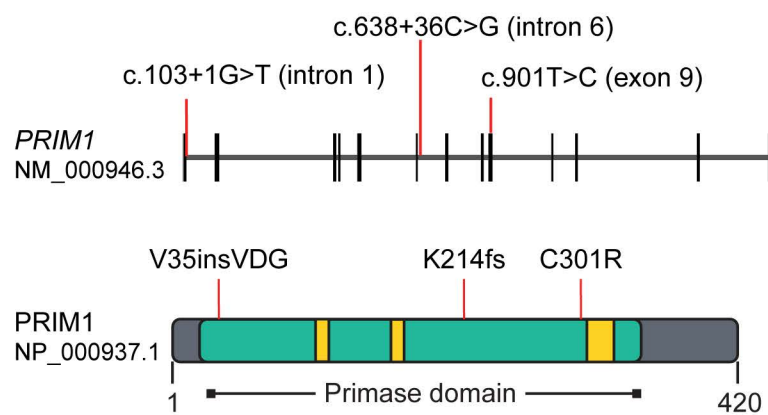
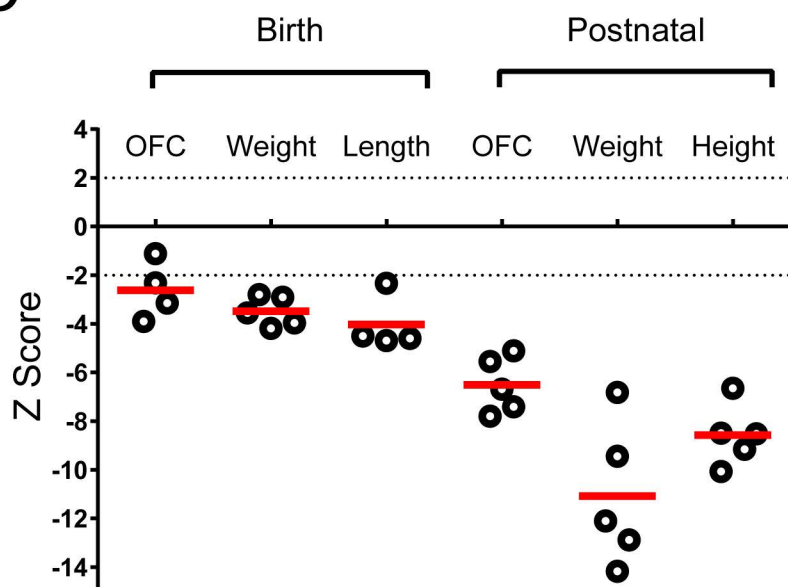
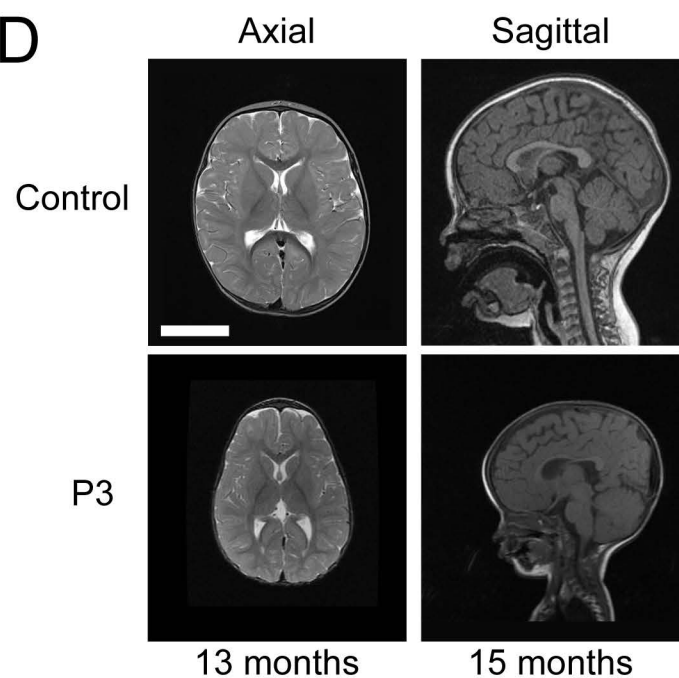
Family	Individual	Allele 1			Allele 2			Country of origin
		HGVSc	HGVSp	AF	HGVSc	HGVSp	AF	
1	P1	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	Hungary
2	P2	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	Hungary
3	P3	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	Hungary
3	P4	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	c.638+36C>G	p.Lys214Serfs*21	1.49 x 10 <sup>-5</sup>	Hungary
4	P5	c.103+1G>T	p.Val35_Ile36insAspGlyVal	2.03 x 10 <sup>-5</sup>	c.901T>C	p.Cys301Arg	-	UK

**Table 1: *PRIM1* Variants identified in individuals with MPD.** Variants reported in this study are described using HGVS nomenclature (<https://varnomen.hgvs.org/>) for the reference coding DNA (HGVSc, NCBI Reference Sequence NM\_000946.3) and protein (HGVSp, NCBI Reference Sequence NP\_000937.1) sequences. The variant allele frequency (AF) for non-Finnish European populations in gnomAD is given for each variant. DNA variants are expressed relative to the coding (“c.”) sequence, and all protein changes are preceded by “p”.

Individual	P1	P2	P3	P4	P5
Family	F1	F2	F3	F3	F4
Gender	Female	Female	Male	Male	Male
Age	Died at 7 mo	7 yrs	Died at 2 yrs	Died at 19 mo	Died age 2 yrs
Consanguinity	y	n/a <sup>b</sup>	n	n	n
IUGR, microcephaly & short stature <sup>c</sup>	y	y	y	y	y
Recurrent infections	Klebsiella pneumoniae <sup>g</sup> ; Coag. neg. Staph <sup>g</sup> . Died of RSV + parainfluenza	Ventriculitis (post-shunt); Rotavirus, Clostridium difficile; Proteus mirabilis <sup>g</sup> .	Respiratory + GI. Recurrent fevers. Died of Adenovirus infection.	Respiratory + GI. Rotavirus enteritis. Enteropathogenic E.coli.	Lower respiratory tract infections. Died of norovirus GI infection.
Feeding issues	y (temp n/g tube)	y	y (temp n/g tube)	y (temp n/g tube)	y (n/g tube)
Developmental Delay <sup>f</sup>	Severe	Severe	Severe	Severe	Gross motor delay
Eyes	Blepharophimosis, small eyes	Blepharophimosis, small eyes	Blepharophimosis, small eyes	Blepharophimosis, small eyes, cataract	Blepharophimosis, microphthalmia
low set ears, microtia	y	y	y	y	y
Other facial features	Broad forehead, late closing fontanelle, flat nasal bridge	Scaphocephaly, small teeth	Frontal bossing, broad forehead, small mandible. Prominent lips.	Prominent lips	Frontal bossing, sparse hair, depressed nasal tip, smooth philtrum.
Absent subcutaneous fat	y	y	y	y	y
Skeletal	Tapering fingers. Small hands and feet, 11 pairs of ribs.	Craniosynostosis. Small hands and feet, tapering fingers.	Tapering fingers. Small hands and feet.	Tapering fingers. Small hands and feet. Late closing fontanelle.	Tapering fingers. Small hands and feet. Late closing fontanelle. Hip dysplasia.
Haematology	Lymphopenia. Thrombocytopenia and anemia <sup>d</sup>	Episode thrombocytosis.	Lymphopenia. Recurrent thrombocytopenia and episode anemia.	Lymphopenia. Thrombocytopenia and anemia <sup>d</sup>	Lymphopenia. Episode thrombocytopenia.
Immunology	IgM, IgA, IgG undetectable	Low IgG	Low IgA	Low IgG. B cell aplasia	Low IgM, IgA, IgG, IgE. B cell aplasia
Cardiac	PDA, PFO	PFO	Mild hypertrophic myocardium	ASD. Pulmonary hypertension	
Respiratory	Chronic wheeze and cough, progressive airway disease, died ARDS	Mild respiratory problems. CXR: increased perihilar density	Chronic lung disease. CXR: infiltrates/consolidation. ARDS.	Chronic respiratory insufficiency. Oxygen dependent. CXR: infiltrates	Episode of severe pneumonitis. Diffuse parenchymal lung disease
Hepatic	Elevated AlkP	-	Elevated LFTs. Macronodular cirrhosis.	Elevated LFTs Hepatic Fibrosis	Elevated AlkP. Nodular hyperplasia. Portal hypertension
Endocrine	-	n/a	Hypothyroid. GH deficient.	Hypothyroid	Hypothyroid. Pancreatic insufficiency
other	-	-	Cryptorchidism	Cryptorchidism	Thymic atrophy. Accessory spleen. Small kidneys. Eczema, severe napkin dermatitis. Low level mosaic, ring chromosome <sup>e</sup>

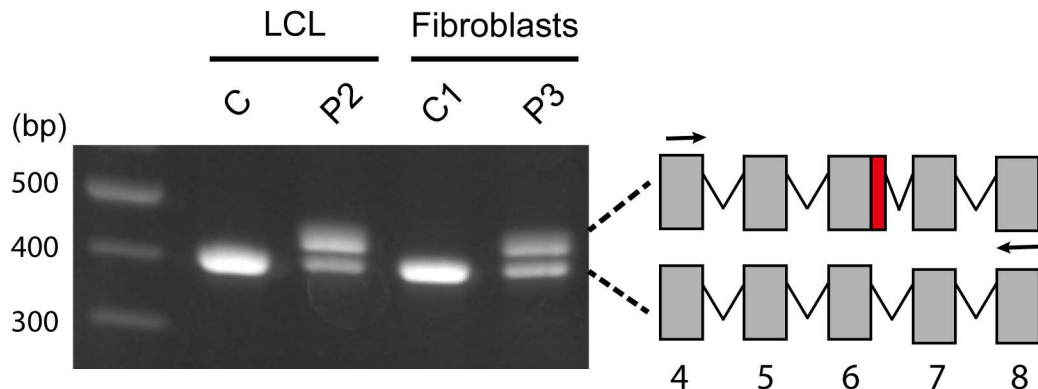
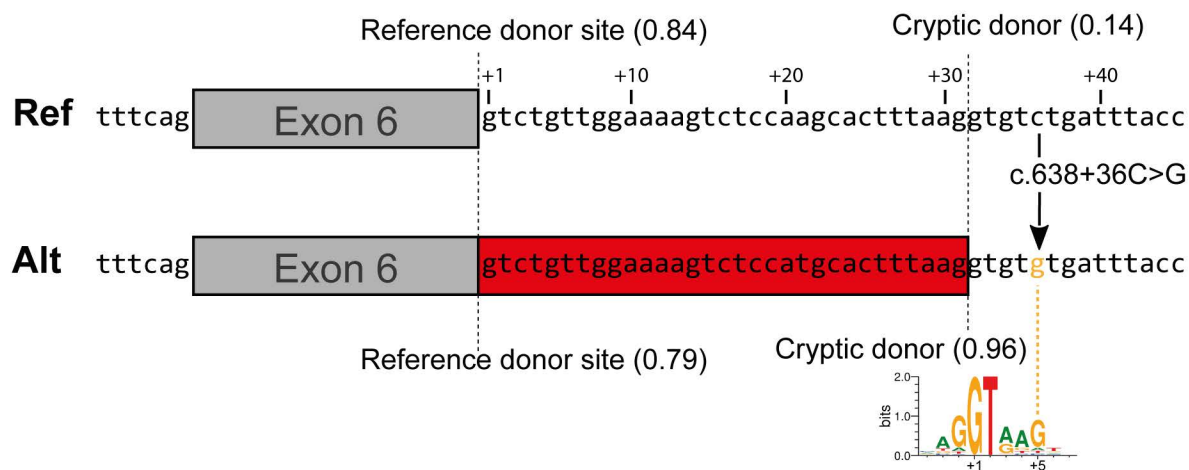
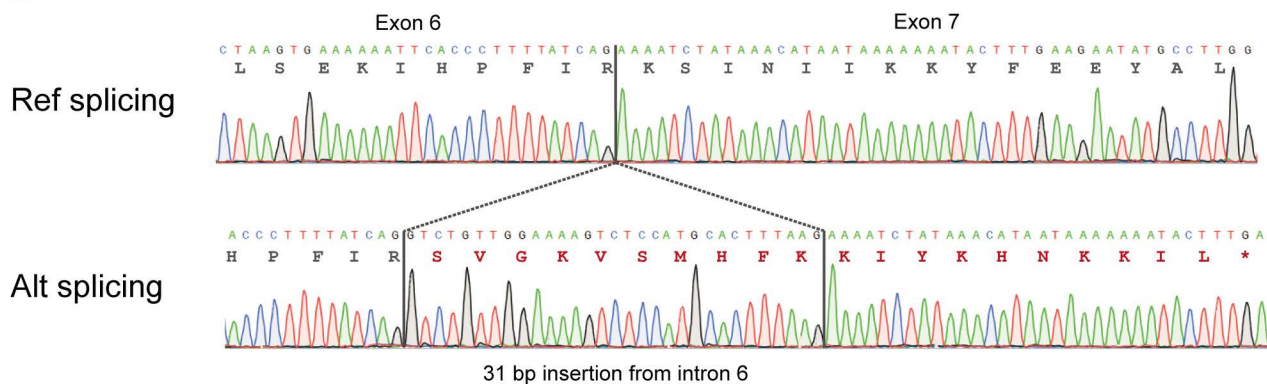
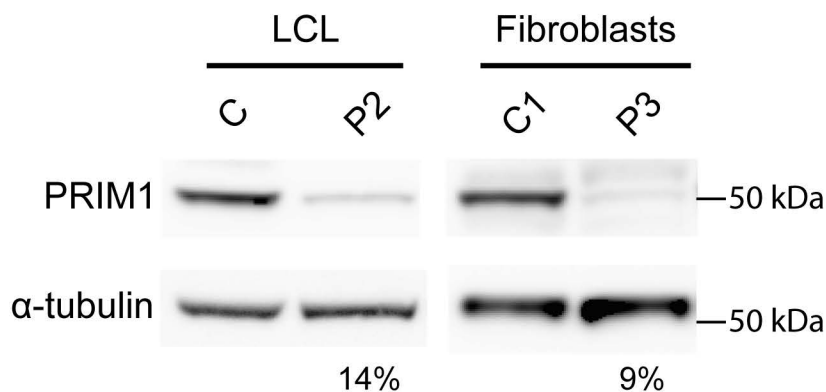
Table 2. Phenotype and clinical features of *PRIM1* subjects.

mo, months; yrs, years; n/a, not available; GI, gastro-intestinal; Coag. neg Staph, coagulase negative Staphylococcus; RSV, respiratory syncytial virus, n/g, nasogastric; Ig, immunoglobulin; PDA, patent ductus arteriosus; PFO, patent foramen ovale; ASD, atrial septal defect; ARDS, acute respiratory distress syndrome; CXR, chest x-ray; LFTs, liver function tests; AlkP, alkaline phosphatase. <sup>a</sup>Likely older affected male sib, IUGR, description similar. <sup>b</sup>Limited family history available. <sup>c</sup>Extreme microcephaly/short stature, defined as occipital-frontal head circumference (OFC) and height respectively  $\leq -4$ sd from age/sex-adjusted population mean. <sup>d</sup>Anemia precipitated by infection. <sup>e</sup>Ring chromosome 20 in 3/30 cells blood only; not detected skin or buccal. <sup>f</sup>Motor greater than intellectual delay – latter assessed as moderate in P1-4. <sup>g</sup>blood culture.

**A****B****C****D****E**

**Figure 1. Individuals with Biallelic *PRIM1* Variants have Primordial Dwarfism.**

- (A) Family pedigrees with segregation of *PRIM1* variants as indicated. Square, male; circle female; filled symbols, individuals with primordial dwarfism.
- (B) Schematic of *PRIM1* transcript and protein. Vertical black lines, exons; location of variants indicated by red lines. Primase domain (NCBI CDD cdo4860), green; nucleotide binding residues, yellow.
- (C) Growth parameters of individuals with *PRIM1* deficiency. Z scores (standard deviations from population mean for age and sex). Dashed lines, 95% confidence interval for general population. Circles, individual subject data points; red bars, mean values.
- (D) MRI neuroimaging of P3 demonstrates microcephaly with simplified gyri. Axial T2; Sagittal T1. Comparison with age matched healthy controls. Scale bar, 5 cm.
- (E) Photographs of individuals P2-P5. P2, 8 months, P3, 7 months; P4, newborn; P5, 7 months; . Written consent obtained from families for photography.

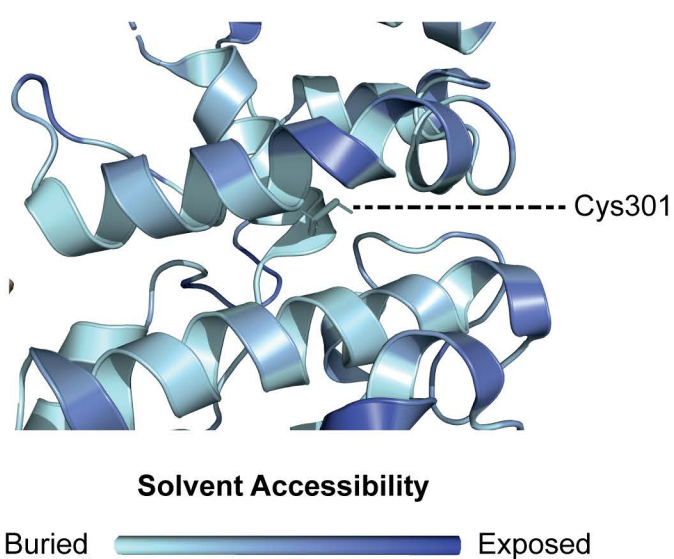
**A****B****C****D**



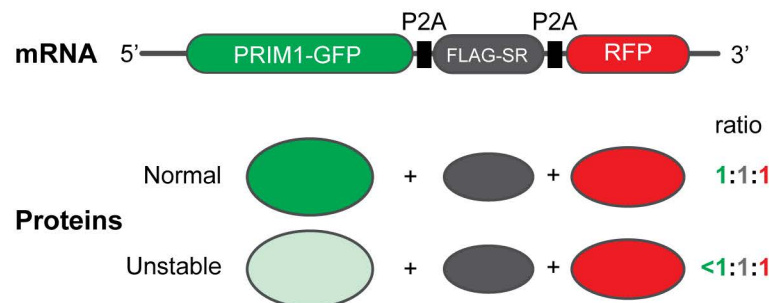
**Figure 2. The c.638+36C>G Variant Creates a Cryptic Splice Donor Site, Resulting in Markedly Reduced Cellular PRIM1 Protein Levels.**

- (A) The c.638+36C>G variant alters *PRIM1* splicing by creating a cryptic splice donor site in intron 6. RT-PCR of *PRIM1* transcripts from lymphoblastoid cells from P2 and primary fibroblasts from P3. Arrows in schematic indicate the position of primers. Grey boxes, exons. Red box, retained sequence resulting from mis-splicing.
- (B) Schematic depicting the effect of the c.638+36C>G variant on splicing of intron 6 of *PRIM1*. The reference and alternate sequences of intron 6 shown with positions of reference and cryptic splice donor sites marked by dotted lines. SpliceAI scores for the donor sites in brackets. 31 nucleotides of included as a result of the c.638+36C>G variant are shown on a red background. A sequence logo created with 100,000 randomly selected human U2 splice donor sites from Ensembl (v83) (Cunningham et al. 2015) using WebLogo (Crooks et al. 2004) illustrates how the c.638+36C>G substitution creates a strong splice consensus sequence by providing a G at the +5 position.
- (C) Representative Sanger sequencing traces of splice products relating respectively to the lower band from panel A ('ref splicing') and the alternatively spliced upper band ('alt splicing').
- (D) PRIM1 protein levels are markedly reduced in cells from individuals P2 and P3, homozygous for the c.638+36C>G variant. Immunoblots of total cell extracts from lymphoblastoid cells (P2) and primary fibroblasts (P3).  $\alpha$ -tubulin, loading control. C, lymphoblastoid, C1, fibroblast cell lines from control subjects. Quantification of PRIM1 protein levels for P2 and P3 cells relative to C and C1 controls respectively (normalized to  $\alpha$ -tubulin loading control) stated beneath each blot.

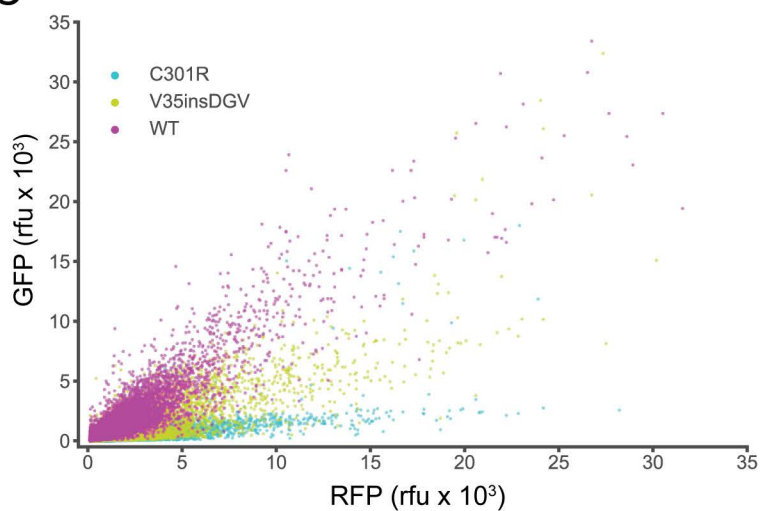
A



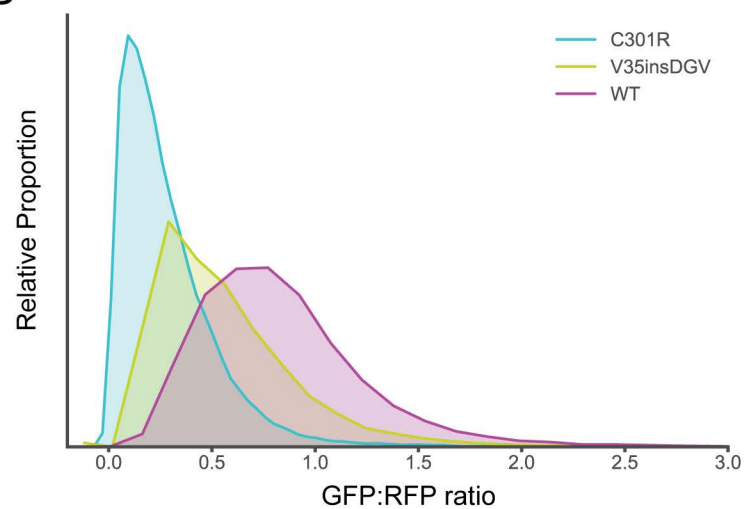
B



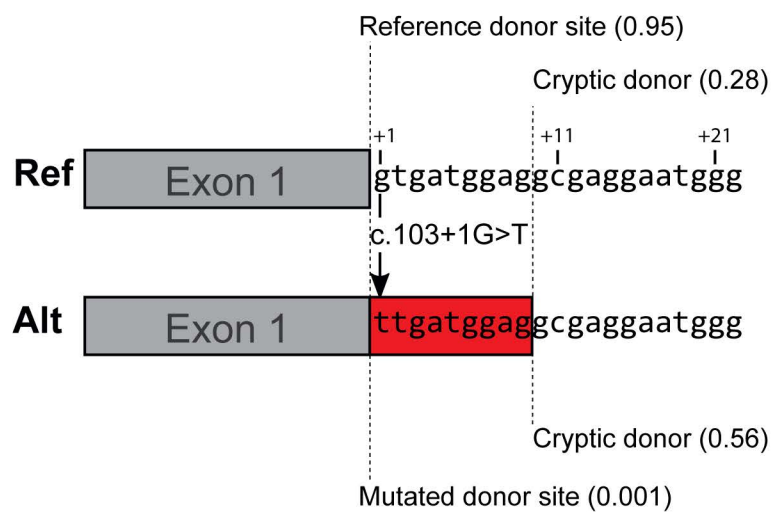
C



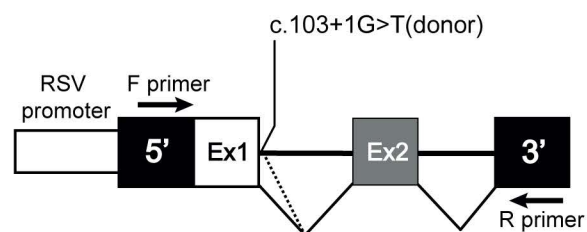
D



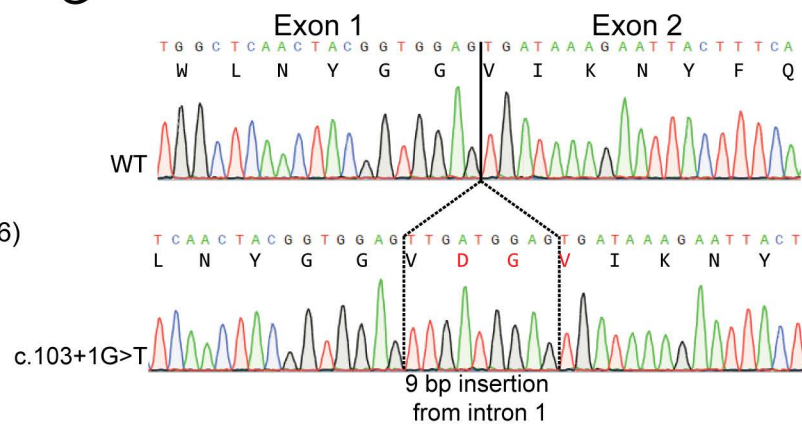
E



F



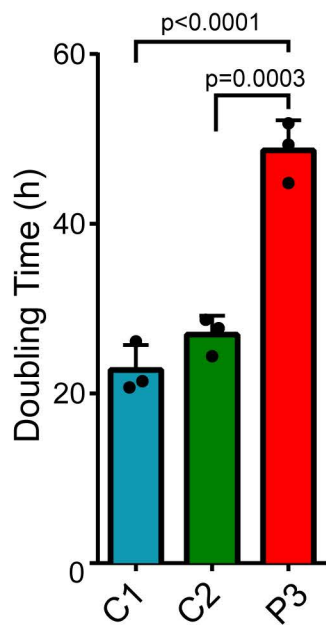
G



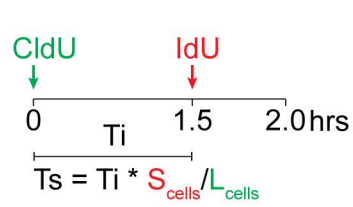
### Figure 3. The c.103+1G>T and C301R variants reduce PRIM1 protein levels

- (A) Cysteine 301, substituted to arginine in P5, lies in a buried hydrophobic region. DNA primase dimer crystal structure (PDB: 4BPU) with residues shaded according to solvent accessibility.
- (B) Schematic of the FACS-based dual reporter stability assay. Expression vector expresses an mRNA encoding PRIM1-GFP-P2A-FLAG-SR-P2A-RFP. Intervening P2A “self-cleaving” peptide sequences produce PRIM1-EGFP, FLAG-SR and RFP polypeptides in equimolar amounts. PRIM1-GFP (wildtype and mutants) and RFP levels are assayed in individual cells by flow cytometry. PRIM1-GFP and FLAG-SR levels can be independently assessed by immunoblotting (See Supplemental Figure S3D).
- (C) GFP-RFP scatter plot for wild type, C301R and V35insDGV dual reporter constructs. n= WT, 27,027; C301R, 44,863; V35insDGV 46,441 cells respectively. rfu, relative fluorescence units.
- (D) Kernel density estimation plot of GFP:RFP ratios from (C).
- (E) Schematic depicting the consequence of the c.103+1G>T variant on splicing. Reference and alternate sequences of intron 1 are shown with positions of the reference and cryptic splice donor sites marked by dotted lines. SpliceAI scores for splice donor sites in brackets. Red box, 9 nucleotides of intron 6 included as a result of activation of the cryptic splice donor variant.
- (F) Schematic assaying the effect of the c.103+1G>T variant. Minigene assay. DNA spanning exon1 (Ex1) and exon 2 (Ex2) of *PRIM1* was cloned into the minigene. Arrows, position of PCR primers. Dotted line, splicing from cryptic splice donor.
- (G) Representative cDNA Sanger sequence traces from wild-type (WT) and c.103+1G>T variant minigene constructs. The three amino acid (DGV) insertion from the c.103+1G>T variant marked in red.

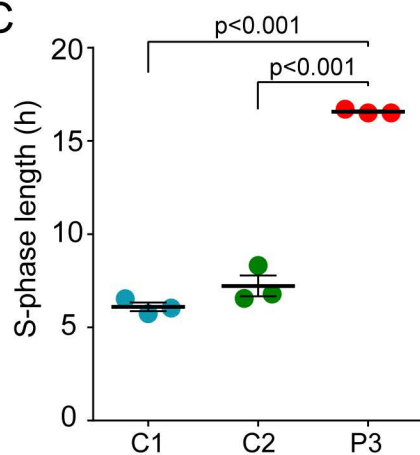
A



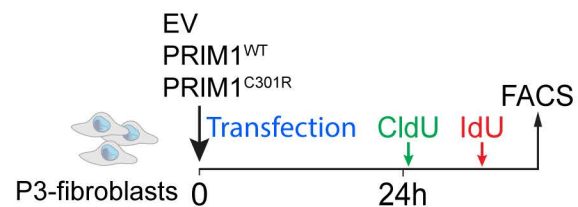
B



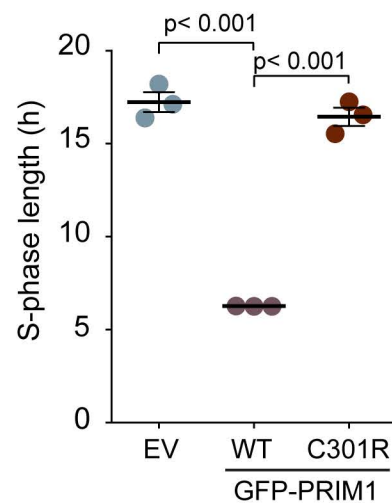
C



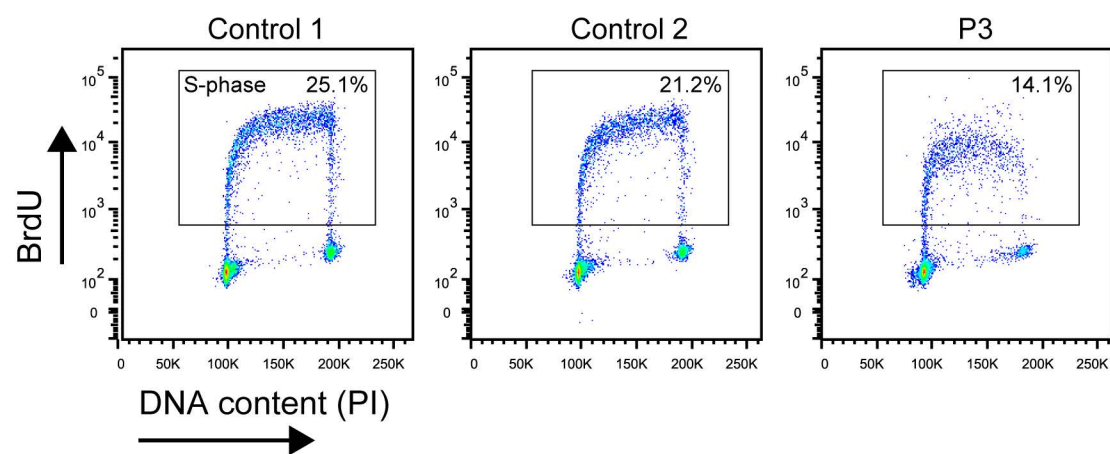
D



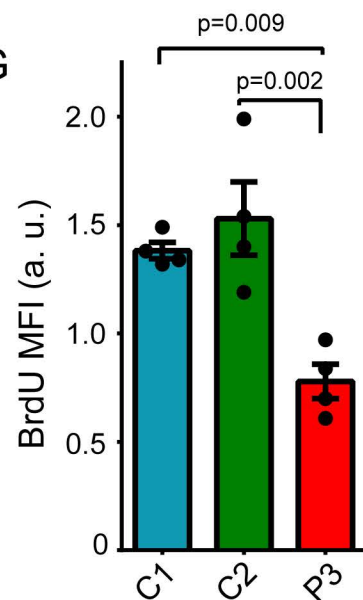
E



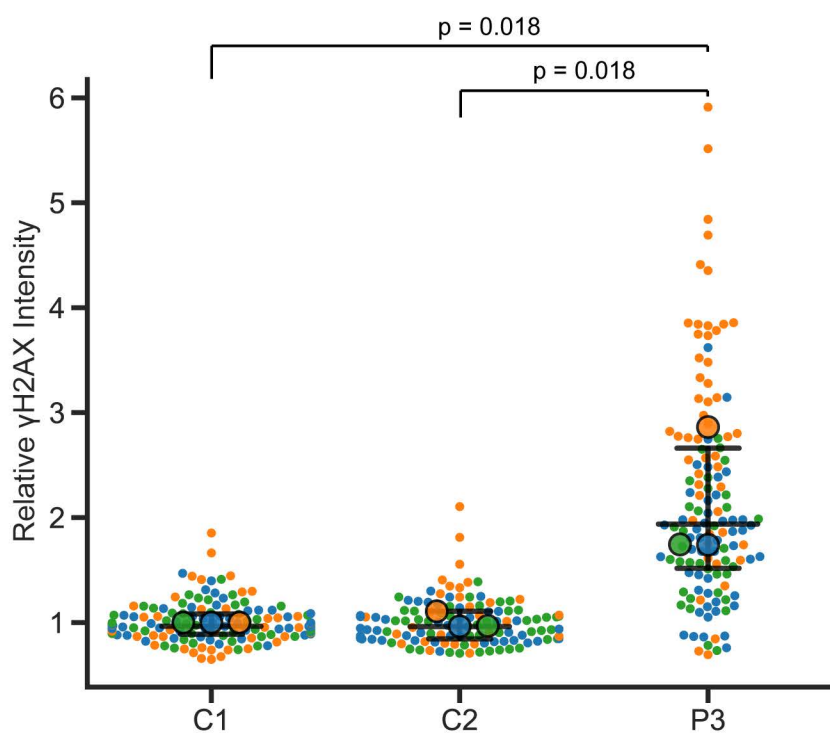
F



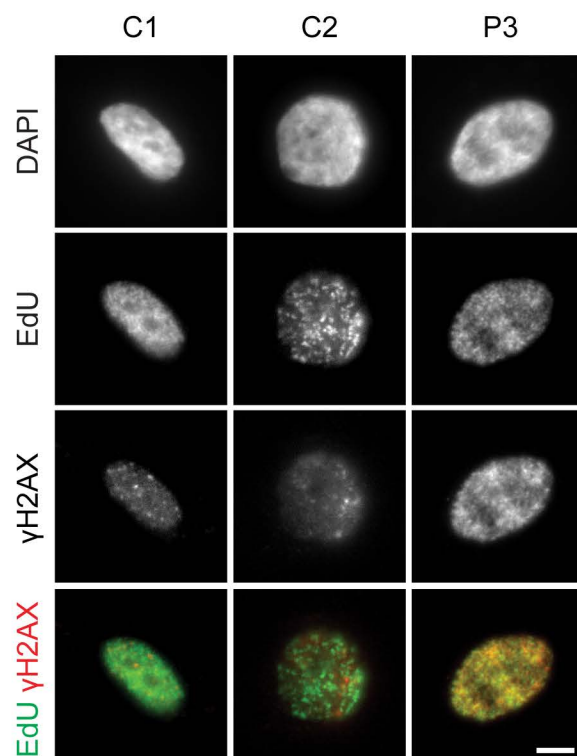
G



H

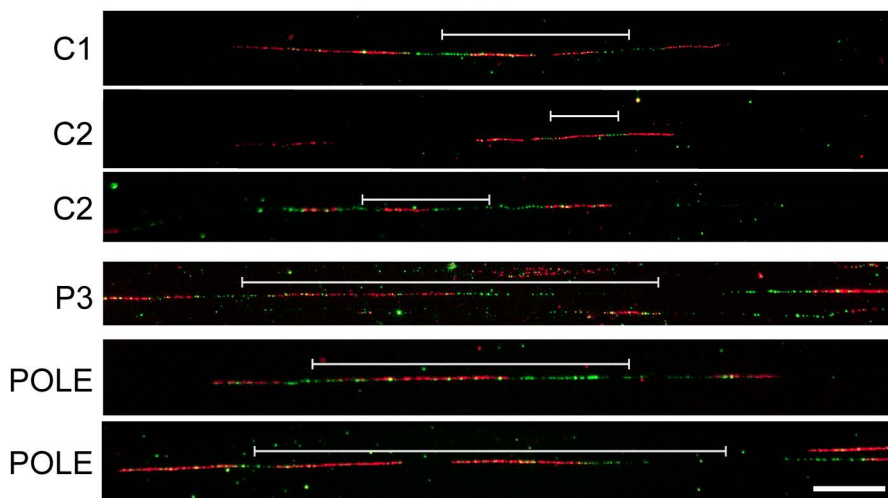
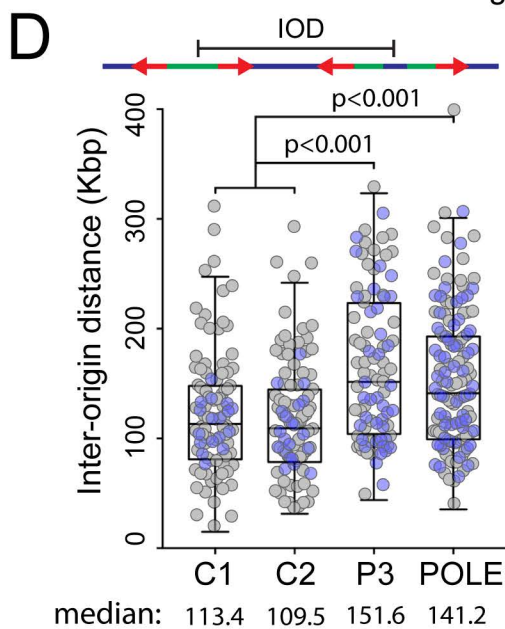
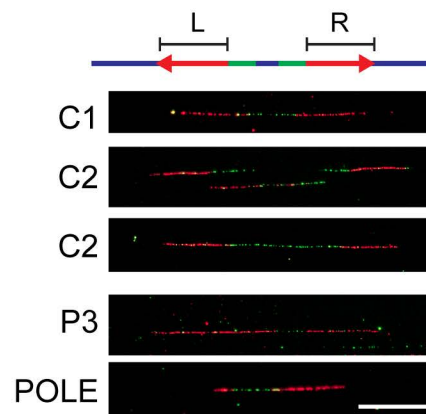
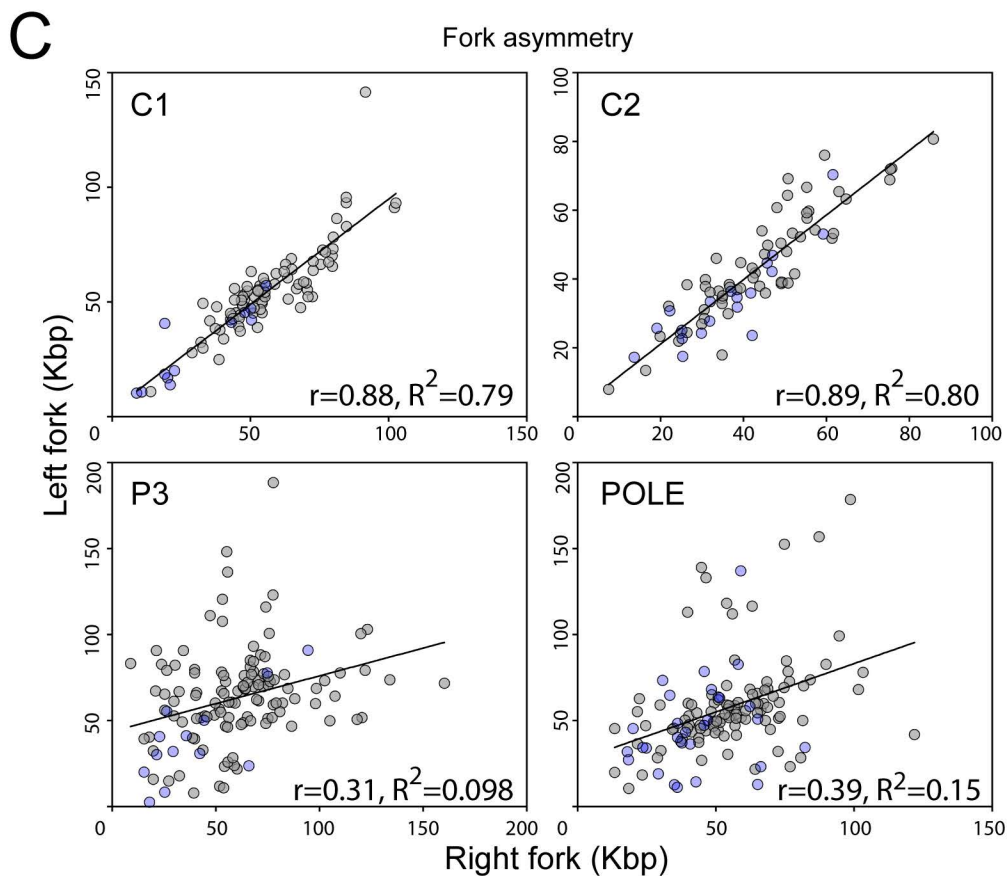
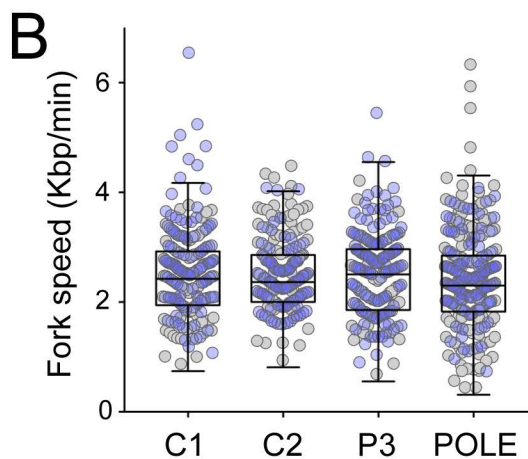
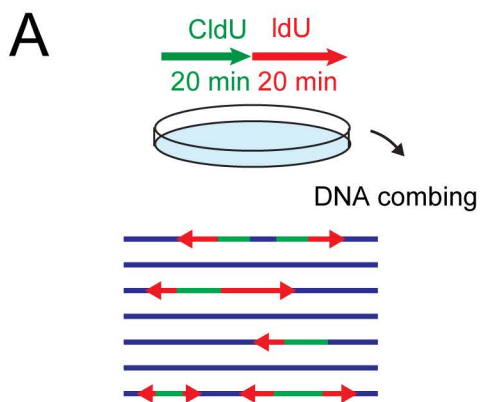


I



**Figure 4. Reduced Cell Proliferation and Impaired DNA Replication in PRIM1 Deficient Primary Fibroblasts.**

- (A) Cell doubling time plotted for 3 independent expts on P3 and two unrelated control C1 and C2 primary fibroblast cell lines. Bars, mean. Error bars, SD.
- (B) Schematic of CldU/IdU double pulse experiment used to determine S-phase time. Cells were labelled with CldU at  $t = 0$  followed by IdU after 1.5 hours. Cells leaving S-phase ( $L_{cells}$ ) are labeled with CldU only, while cells remaining in S-phase ( $S_{cells}$ ) are labeled with both CldU and IdU.  $T_s$ , S-phase length, the product of interval between pulses ( $T_i$ ) and the proportion of  $S_{cells}$  to  $L_{cells}$  (Martynoga et al. 2005).
- (C) S-phase time is substantially increased in P3 fibroblasts compared to controls. Mean $\pm$ SEM, N=3 expts.
- (D) Schematic of rescue experiment. P3 fibroblast transfected with either empty vector (EV), wild-type (WT) or C301R-GFP-PRIM1 as indicated and after 24 h S-phase time determined as in (B) for GFP<sup>+</sup> cells.
- (E) Complementation with WT GFP-PRIM1 rescues slow S-phase progression in P3 fibroblasts. S-phase length plotted for n=3 expts. Mean $\pm$ SEM.
- (F) DNA content and BrdU flow cytometry scatter plots, representative of 4 independent experiments on control (C1, C2) and P3 primary fibroblast cell lines.
- (G) BrdU incorporation is reduced in PRIM1 deficient cells during S-phase. Quantification of BrdU mean fluorescence intensity (MFI) from control and patient-derived fibroblasts according to S-phase gate in (F); a. u., arbitrary units.
- (H)  $\gamma$ -H2AX is increased in S-phase P3 fibroblasts. Mean  $\gamma$ -H2AX intensity calculated for EdU-positive nuclei from C1, C2 and P3 cells. n=3 experiments. Data points coloured by experiment; filled circles, mean values for each replicate; bars, median and interquartile range (all values); values normalised for each experiment relative to C1 mean value. P-values, repeat-measures ANOVA with Tukey multiple comparison test.
- (I) Representative immunofluorescence images of S-phase nuclei quantified in (H). Scale bar, 5  $\mu$ m. Statistics in A, C, E, and G, one-way ANOVA with Tukey multiple comparison test.



**Figure 5. PRIM1 Deficiency Causes Fork Asymmetry and Increased Inter Origin Distances, Consistent with Replication Fork Stalling and Decreased Replication Initiation.**

- (A) Schematic depicting DNA combing experiments. Sequential 20 minute CldU and IdU pulses were performed on cultured primary fibroblasts, which were then harvested, with DNA combing performed to characterise DNA replication at the single molecule level.
- (B) DNA fork speed, Kbp/min, in primary fibroblasts from two unrelated control individuals (C1, C2), a PRIM1-deficient individual (P3), and a POLE-deficient individual. Grey and blue dots represent measurements from (n=2) respective independent experiments (B-D).
- (C) PRIM1 deficient cells exhibit fork asymmetry, similar to POLE deficient cells. Left (L) vs. right (R) fork ratio scatter plots.  $r$ , Pearson correlation coefficient. Right, representative images of bi-directional forks. Scale bar, 20  $\mu$ m.
- (D) Inter origin distance (IOD) in control, PRIM1-, and POLE-deficient primary fibroblasts; boxplots graphed according to Tukey method (B,D). P-values, Kruskal-Wallis test. Right, representative images of IODs. Scale bar, 20  $\mu$ m.